

Epigenomics project

ChIP-Seq

Target TF: USF1

Data: <https://www.encodeproject.org/experiments/ENCSR000BKT/>



Sabrina Pervin Abu

Upstream Transcription Factor 1

Function:

USF1 is ubiquitous transcription factor.
The protein can activate transcription through E-box motifs.
It regulates the expression of many genes involved in lipid metabolism, immune response, endothelial function and aging.

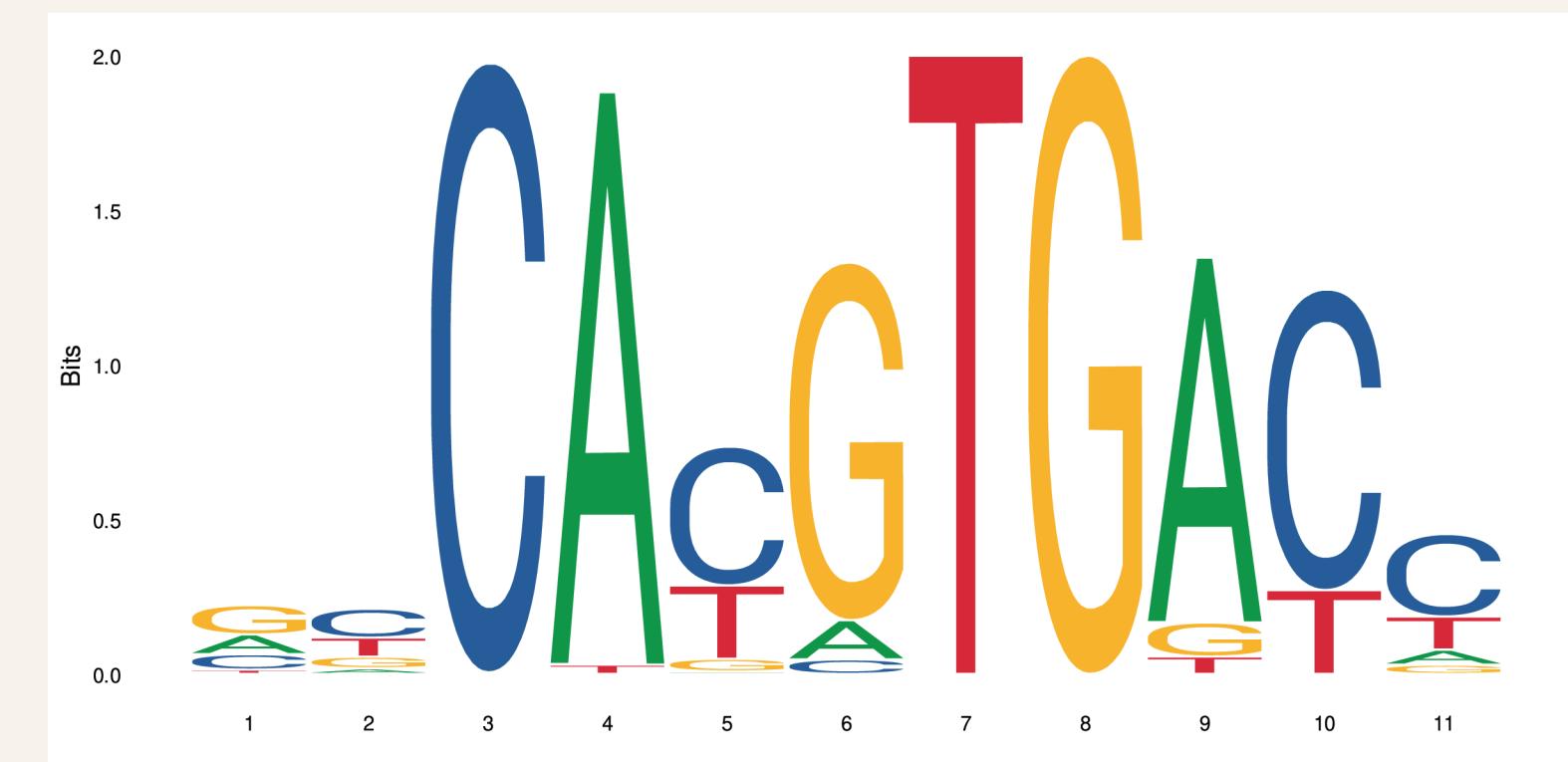
Fan, YM., Hernesniemi, J., Oksala, N. et al. Upstream Transcription Factor 1 (USF1) allelic variants regulate lipoprotein metabolism in women and USF1 expression in atherosclerotic plaque. Sci Rep 4, 4650 (2014). <https://doi.org/10.1038/srep04650>

DNA binding domain:

It is member of the basic helix-loop-helix leucine zipper family.
(bHLH-Zip)

Binding motif:

E-box sequences (5'-CACGTG-3')



<https://jaspar.elixir.no/matrix/MA0093.2/>

Preliminary mapping and quality controls

Sample	Total Reads	Uniquely Mapping Reads	Non Mapping Reads	Multi Mapping Reads
Replicate1	24,312,794	21,545,349 (88.62%)	2,654,909 (10.92%)	112,536 (0.46%)
Replicate 2	13,319,737	11,646,746 (87.42%)	1,591,174 (11.95%)	81,817 (0.61%)
Control	21,795,072	20,709,511 (95.01%)	1,000,709 (4.59%)	84,852 (0.39%)

 **Extremely low read depth** 

The second replicate has an extremely low number of reads mapped on the reference genome.
It would be preferable to work with at least 20million reads.

The uniquely mapping reads for both replicates are high and meet quality standards.

`samtools view -bq 1 unfiltered_file.bam > filtered_unique.bam`

Peak calling

File	Number of peaks (pre filtering)	Number of peaks (post filtering)	Number of removed peaks (blacklist)	Redundant rate	Fragment size
Replicate 1	43702	43682	20	0.35	121 bp
Replicate 2	20370	20291	79	0.47	158 bp
Merged .bam file	37805	37783	22	0.40	128 bp

Redundant rate: redundant rates are **NOT** below 20%.
The estimated fragment sizes are between 100 and 200 bp

`macs2 callpeak -t rep1_filtered_unique.bam -c control_filtered.bam -g hs -n rep1`

Peak Overlap Analysis

Overlap	Number of overlaps	Overlapping peaks (%)
Simple Overlap Between Filtered Replicates	17271	85,11% [vs REP2] 39,53% [vs REP1]
Summit Proximity Between Replicates (peak summits within 100 bp from one another)	16664	81,80% [vs REP2] 38,14% [vs REP1]
Replicate 1 / Merged Set (both filtered)	35315	93,46%
Replicate 2 / Merged Set (both filtered)	18440	48,80%

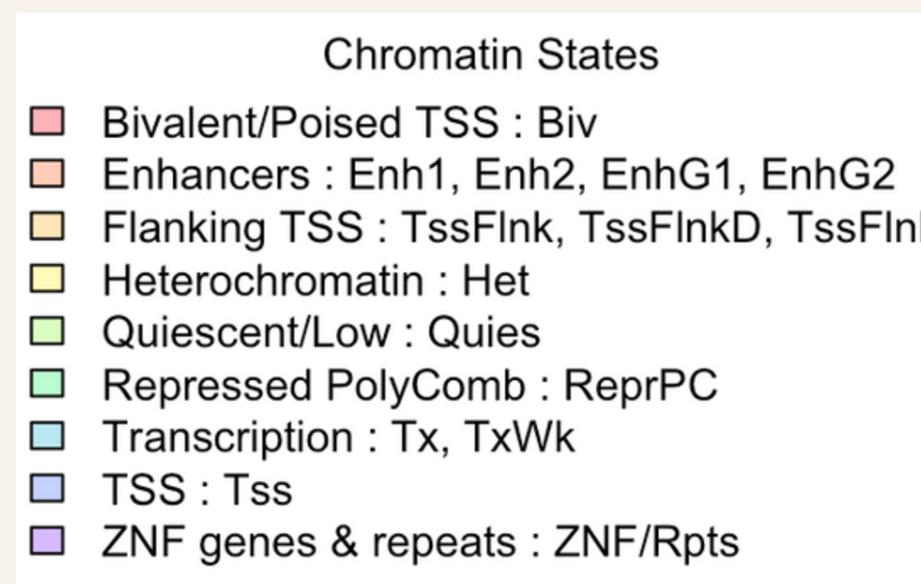
At least 70-75% of the peaks should be in common across replicates: the requirement is not met due to the great difference in reading depth between replicates.

Comparison with ENCODE

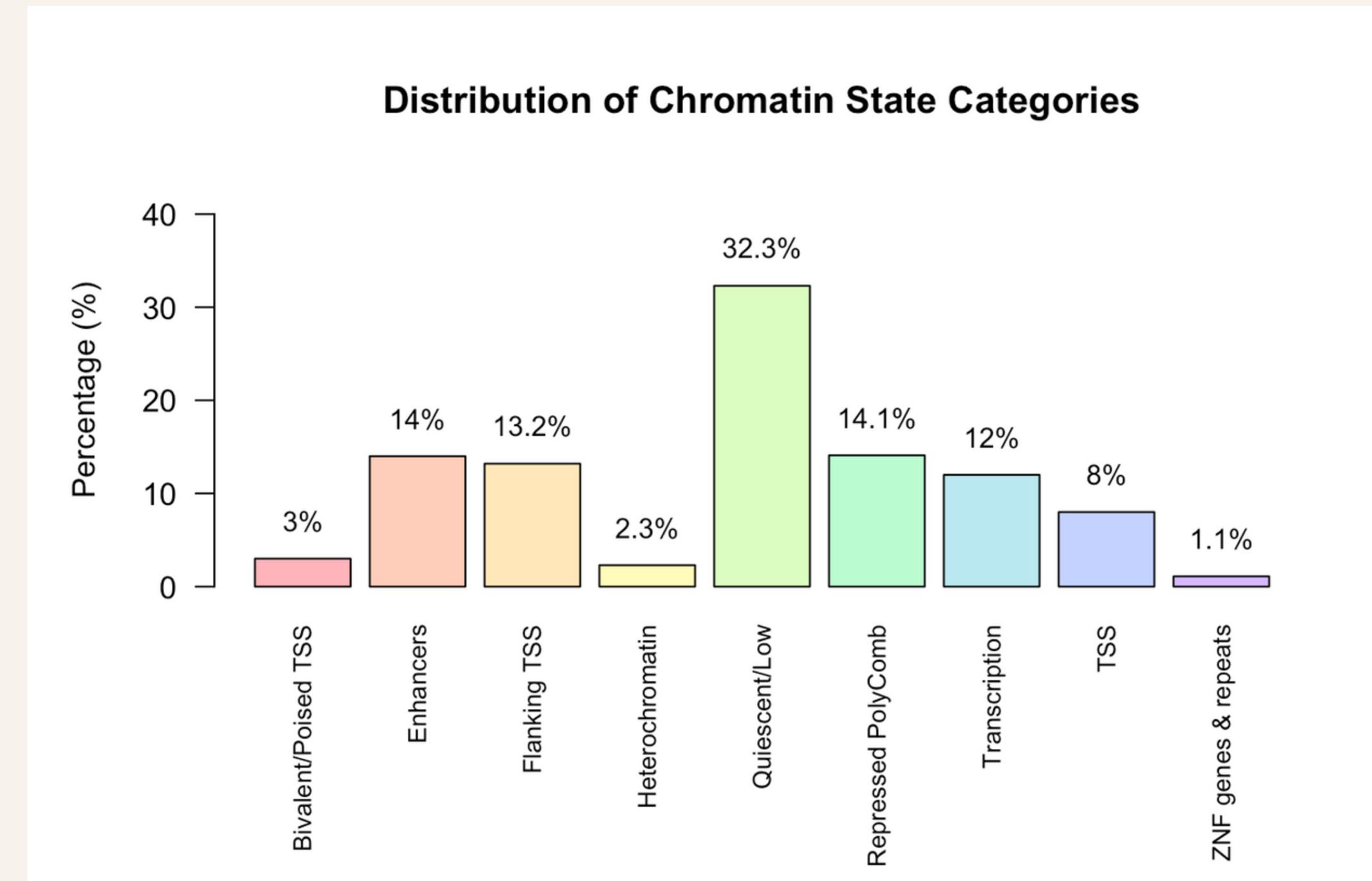
Comparison	Overlapping Peaks	Jaccard Index
Replicate 1	21154	0.40
Replicate 2	13837	0.35
Intersection Set	13726	0.38
Merged Set	21369	0.44

Final peak-set

Chromatin States associated with USF1 binding



Most of the peaks are in **Quies**



Target Genes

GREAT was used to find the genes regulated by USF1, by defining:

- Promoter: +/- 1 kb around the TSS
- Distal region: 30 kb from the TSS

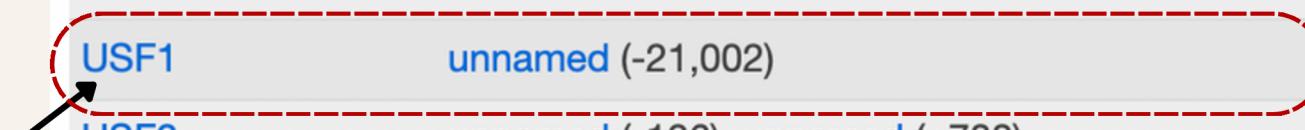
Number of target genes because USF1 binds to their promoter: 5305

Number of target genes because USF1 binds to a distal regulatory region: 9520

Does USF1 regulate itself? Yes

Gene -> genomic region association table [Download table as text.](#)

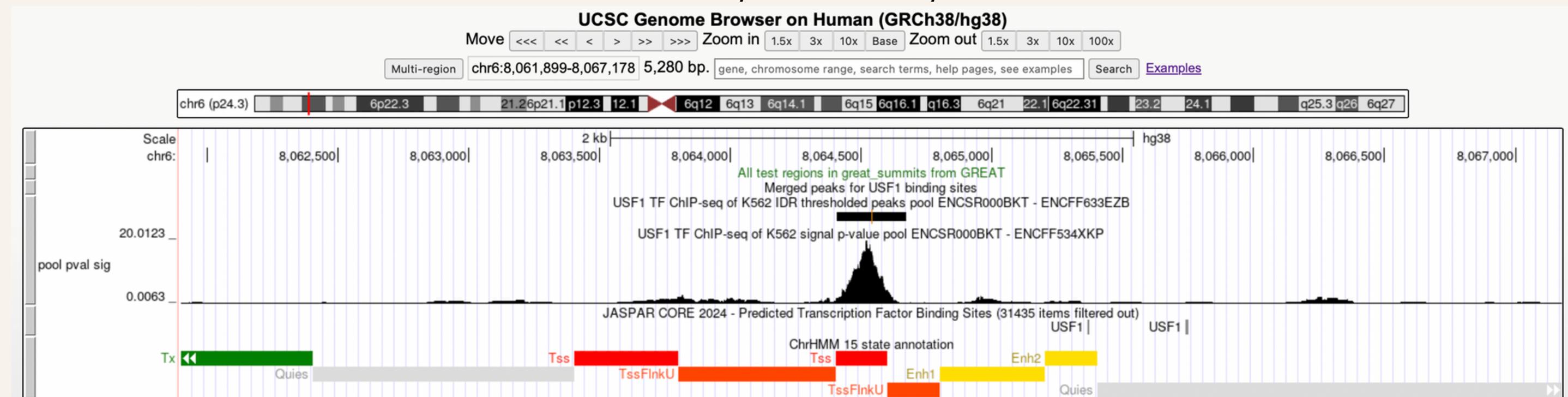
UROD	unnamed (-765), unnamed (-8), unnamed (+590), unnamed (+27,089)
USB1	unnamed (-95)
USE1	unnamed (-34), unnamed (+10,956), unnamed (+11,884), unnamed (+15,831)
USF1	unnamed (-21,002)
USF2	unnamed (-126), unnamed (+788)
USH1C	unnamed (+15,825)
USH1G	unnamed (+21,347)
USHBP1	unnamed (-28), unnamed (+15,292)
USO1	unnamed (-160), unnamed (+21,724)



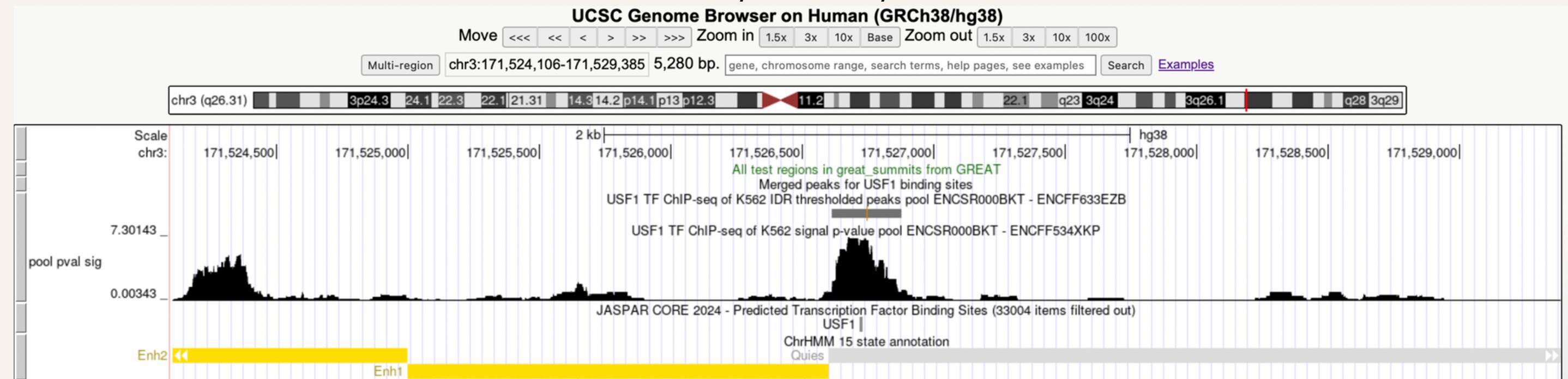
USF1

Visualization (UCSC Genome Browser)

Peak found by ENCODE but not by me:



Peak found by me but not by ENCODE:



Does USF1 regulate itself?

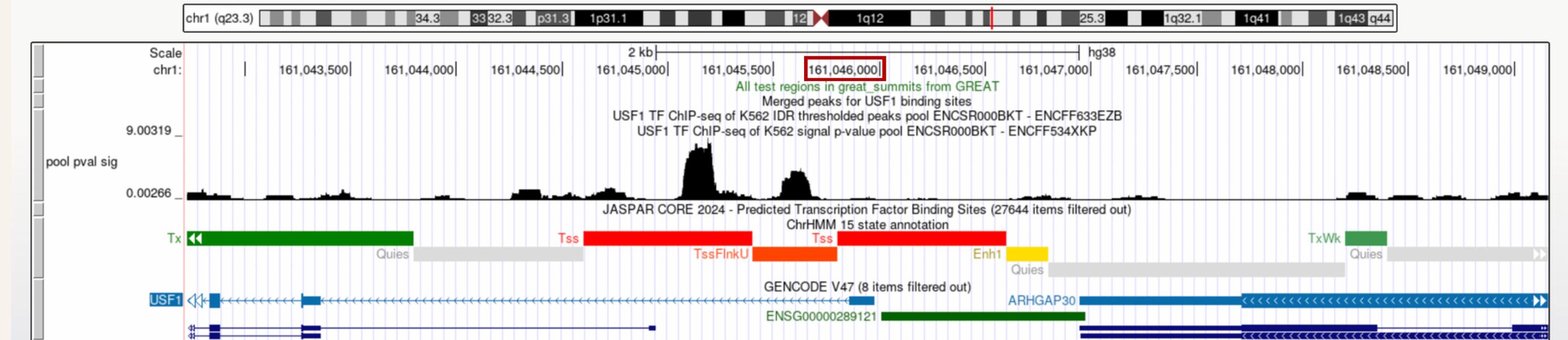
Gene -> genomic region association table [Download table as text.](#)

USF1 unnamed (-21,002)

UCSC Genome Browser on Human (GRCh38/hg38)

Move <<< << < > >>> Zoom in 1.5x 3x 10x Base Zoom out 1.5x 3x 10x 100x

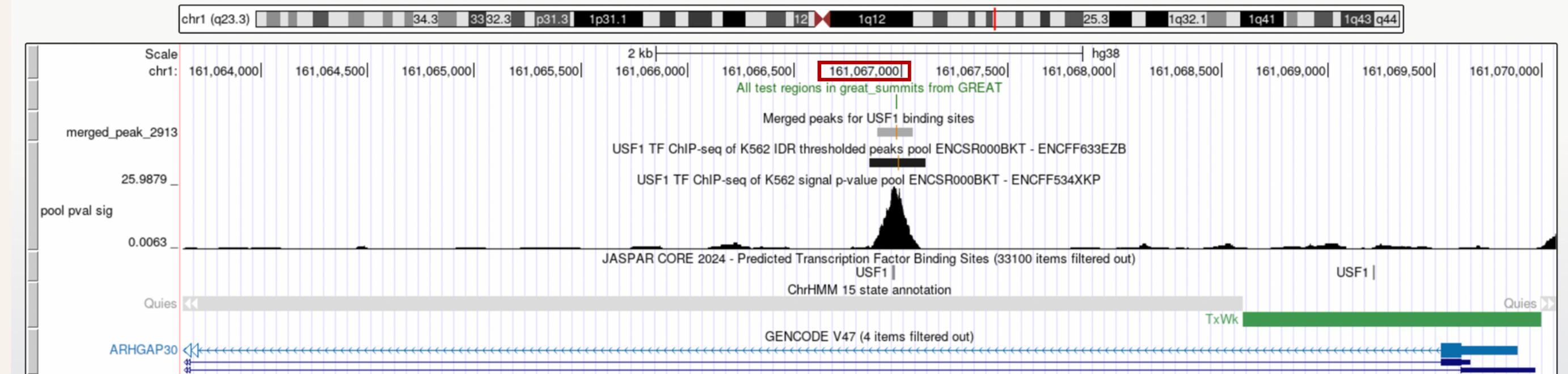
Multi-region chr1:161,042,728-161,049,161 6,434 bp. gene, chromosome range, search terms, help pages, see examples Search Examples



UCSC Genome Browser on Human (GRCh38/hg38)

Move <<< << < > >>> Zoom in 1.5x 3x 10x Base Zoom out 1.5x 3x 10x 100x

Multi-region chr1:161,063,636-161,070,069 6,434 bp. gene, chromosome range, search terms, help pages, see examples Search Examples



Motif Analysis (PscanChIP)

Name	ID	L.PV	L.O/U	G.PV	G.O/U	SP.COR	P.POS	P.POS.PV
MAX::MYC	MA0059.1	0		0		0.0007	[0,10]	0
Pax2	MA0067.1	0		0		-0.0324	[-6,4]	0
NR4A2	MA0160.1	0		0		-0.008	[-6,4]	6.0E-104
ARNT::HIF1A	MA0259.1	0		0		-0.022	[-1,9]	0
NFYB	MA0502.1	0		0		0.0031	[7,17]	1.2E-25
ZBTB33	MA0527.1	0		0		-0.0053	[-3,7]	2.1E-176
ESR2	MA0258.2	0		0		0.0171	[-3,7]	1.7E-134
USF1	MA0093.2	0		0		-0.012	[5,15]	0

Matrix Info

ID	MA0093.2
Name	USF1
Class	Basic helix-loop-helix factors (bHLH)
Species NCBI ID	9606
Inf. Content	13.4
SuperGroup	vertebrates
Protein Acc.	P22415
Type	ChIP-seq
PMID	8052536
Report Best Occurrences	<input type="button" value="Go!"/>

MA0093.2

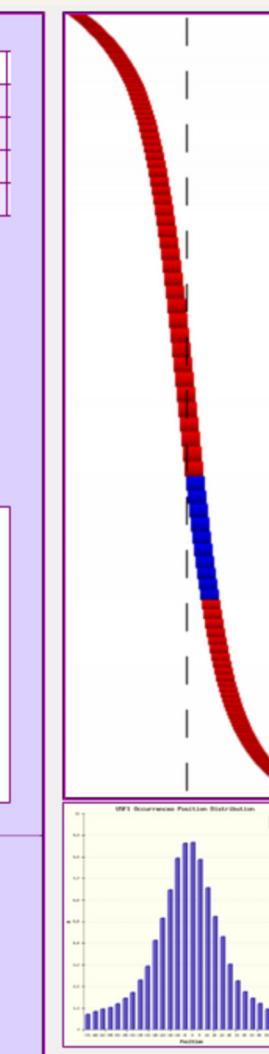
	1	2	3	4	5	6	7
A	4846	1472	43	16573	188	1605	0
C	3764	7448	16799	0	10279	595	0
G	7162	2983	0	0	932	14642	0
T	1070	4939	0	269	5443	0	16842

bits

[Download as txt file](#) [Run Pscan-ChIP centered on these sites](#) [Positional pvalues table](#)

CHR	REG_START	REG_END	REL_SITE_START	REL_SITE_END	SITE_STRAND	SCORE	OLIGO
chrY	13825322	13825471	24	33	+	1	GCCACGTGACC
chrX	153637276	153637425	-20	-11	-	1	GGTCACGTGGC
chrX	153564318	153564467	-33	-24	-	1	GGTCACGTGGC
chrX	153077599	153077748	32	41	-	1	GGTCACGTGGC
chrX	153072172	153072321	66	75	+	1	GCCACGTGACC

Only the top 500 best occurrences reported... download txt file for more occurrences



Matrix Info

ID	MA0067.1
Name	Pax2
Class	Paired box factors
Species NCBI ID	10090
Inf. Content	6.21
SuperGroup	vertebrates
Protein Acc.	P32114
Type	SELEX
PMID	8132558
Report Best Occurrences	<input type="button" value="Go!"/>

MA0067.1

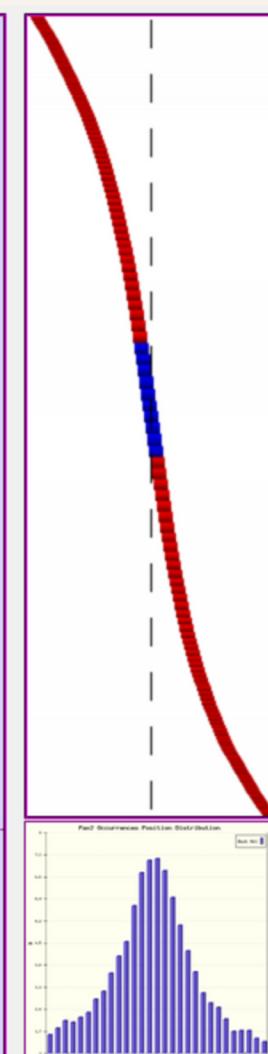
	1	2	3	4	5	6	7	8
A	10	7	3	0	26	2	2	1
C	7	1	2	28	1	17	0	11
G	5	21	0	1	3	1	19	11
T	9	2	26	2	1	11	10	8

bits

[Download as txt file](#) [Run Pscan-ChIP centered on these sites](#) [Positional pvalues table](#)

CHR	REG_START	REG_END	REL_SITE_START	REL_SITE_END	SITE_STRAND	SCORE	OLIGO	
chrX	154504630	154504779	2	14	+	1	AGTCACGC	
chrX	152733664	152733813	2	7	0	1	AGTCACGG	
chrX	152714502	152714651	2	22	+	1	AGTCACGG	
chrX	152699209	152699358	2	54	+	1	AGTCACGG	
chrX	137276149	137276298	2	1	6	+	1	AGTCACGG

Only the top 500 best occurrences reported... download txt file for more occurrences



Thank you!

